INTERACTIVE MULTIMEDIA STATISTICS (IMS)

Alan Jones, Faculty of Health Sciences, University of Sydney, Australia
Susan M. Crowe, Dept. of Statistics, Macquarie University, Australia

IMS is a multimedia software package designed as a set of self directed tutorials to be used as an adjunct to other media in an introductory statistics service course. It is PC based in CD or 1.44Mb disc format and can be used stand alone or on a LAN. With the addition of a plugin it can be accessed in full functionality using a Web browser. Concepts are illustrated using the Minitab 'Bear Data' with cartoon bears, alongside numerical summaries and traditional statistical graphics. The material is structured sequentially to follow the pattern of typical introductory courses. Largely expository in nature but allowing students to discover what we need them to discover interactively.

THE RATIONALE FOR IMS

Teaching introductory statistics in so called service courses can be described as "teaching the unwanted to the unwilling". The challenge for teachers of introductory statistics has been to shift the emphasis from teaching to learning and to have students take control of their learning through avenues such as data exploration and simulation. Often the difficulty is to provide data which students perceive as interesting, relevant to them and which clearly illustrates the concepts we wish them to understand. The difficulties are compounded when courses include students from several substantive areas of study. Existing data sets are often a turn off to students who view them merely as

numbers without context. Collecting “real” data in class, whilst useful, can be unmanageable in large classes, is time consuming and the time lapse between collection and processing fail to give the instant gratification many students want and need to capture the relevance of the exercise. There is also usually little motivation to ‘explore’ data and ‘discover’ the underlying principles. With ever increasing pressure on students’ time there seems to be a growing chorus of ‘... just tell me what I need to pass this course’. Students are becoming greater optimisers with demands for more expository type material as opposed to the more exploratory approach. As teachers of statistics well know, in order to adequately grasp statistical concepts students need to interact with data and extract meaning for themselves.

This was our dilemma when planning PC based interactive material. We sought to blend the expository approach with interactive discovery. An analogy might be a walk in the woods. Rather than simply giving the students a rough map of the woods and telling them to go and discover the wonders of the arboreal world, which did not particularly motivate them in the first place, we lead them through a defined route so that they cannot help discover the wonders we as experienced guides know to be there. This way the walk is shortened and we are assured that they have encountered the material they should discover. IMS is largely expository with we hope enough interaction and discovery material to motivate students.

THE DATA

How many of us have been disheartened by students’ cool or even hostile reaction to the wonderful ‘real world’ or contrived data sets and examples we have dreamed up to motivate our students? What we wanted was a data set which illustrated all introductory concepts, small enough to be manageable, large enough to see patterns and have enough variables to provide some useful insights into the underlying mechanisms ... and we wanted it to be motivational ... and be palatable to students from many disciplines. And so the search for the Holy Grail began... Fortunately our search proved successful. We discovered the ‘Bear Data’ shipped with the later versions of Minitab. The data collected on Black Bear by Gary Alt at Penn State some years ago was a Mecca. There are about 100 records with several variables measured, both categorical, e.g sex, and quantitative, e.g weight, normal and skewed distributions, relationships and much more. Analysis of this data could show some features of the bears and pose some questions, but this was

data, just numbers. In order to bring this data to life we cartooned the bears and used them alongside the numbers. Bears have a special place in the human psyche as cute or cuddly. Using the speed of the PC and programed algorithms we were able to do such things such as present random samples showing the relatively sized bears positioned over a numeric scale with a 'mean bear'(Figure 1) introduced at the click of a button. The

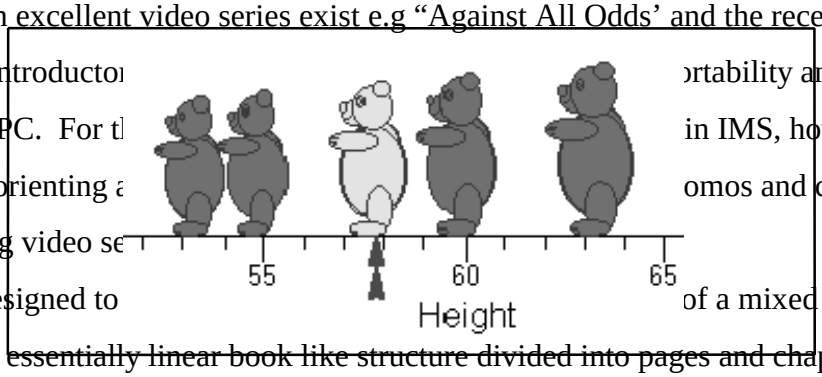

Figure 1. n = 4 with xbar-bear

reaction of students, from diverse disciplines, has been very positive with no complaints that the material is 'irrelevant'.

THE CONTEXT FOR IMS

Multimedia software has potential to deliver material in a new and unique way but it has its' weaknesses if used in the wrong context. Large amounts of text are less comfortable to read from a screen than traditional text media. For this reason we have designed IMS to accompany rather than replace a course text. Many excellent modern texts are available and offer the portability and readability not possible (yet) with a PC. We do acknowledge however that search facility and hyper-linking in PC based text does confer an advantage over traditional text media. We have utilised hyper-linking in IMS but kept text to a minimum.

Video is a useful medium for illustrating context through anecdotes and real world situations. Again excellent video series exist e.g "Against All Odds' and the recent Open University, UK introductor ortability and comfort than on PC. For t in IMS, however we may include orienting a omos and context for accompanying video se

IMS is designed to of a mixed media course. It has an essentially linear book like structure divided into pages and chapters parallelling the sequence used in most standard introductory statistics texts. Our

expository structured approach is quite different from the discovery approach used in the STEPS program from the UK consortium. We think that the two projects are complimentary with the STEPS filling the exploratory role after concepts are introduced. At the core of our mixed media course is a workbook/study guide which directs student learning through reference to a course text, videos and IMS modules and requiring worked problems based on the other material. For on-campus courses we use some lectures to add the quality of human interaction and hopefully motivation. Our approach is to use the inherent strengths of several media rather than a stand alone PC course. We think that this approach gives maximum flexibility in tailoring courses to particular situations such as on campus or distance mode and to students with differing needs.

IMS DESIGN

The overall design of the package is in page and chapter format with hyperlinked tables of contents and indexes. Hyperlinked ‘hotwords’ are used to provide explanations, demonstrations, and glossary information. However it is not ‘a book on the screen’ with large amounts of text. Text has been kept to a minimum and is used only as additional explanation for the screen exercise at hand. We have kept screen layouts simple with few objects and limited colours, a fact that students comment favourably on. Extensive use is made of ‘show and hide’ for e.g posing questions then allowing students button access to explanation or for building up displays sequentially to demonstrate concepts such as the meaning of standard deviations. Extensive use is also made of animation not only to hold the learners attention but to illustrate ideas such as shading areas under curves or moving a 95% confidence band within a population curve.

DELIVERY

The software is PC based and will be made available on 3.5" floppy discs and/or CD making it suitable for LAN or stand alone installation. With the addition of a small plugin full functionality is available on the WWW or on a MAC via a Web Browser.

CONTENT

As mentioned above the modules of the package are complimentary to the sequence and material in most standard introductory statistics service courses. The areas covered are;

1. Introduction to data and data structures in windows based statistical packages with an emphasis on the context of the data showing how the numbers describe particular individuals and their features.
2. Graphical representation of data with an emphasis on interpretation and the discovery of facts about the group bears. Graphs are presented to reveal a profile of the group and pose some further questions which need changes of graph and scale to determine.
3. Numerical summaries. Emphasis is placed on how the numbers relate to the characteristics of the group and to the graphical representations. The idea of describing variation is introduced at this stage with reference to patterns and departure from the pattern.
4. Distributions are introduced with left skewed histograms of age which are developed in to density distributions. Normal distributions are also introduced in this way using anatomical variables. In both cases emphasis is placed on understanding the underlying mechanisms e.g life is hard in the woods so fewer and fewer bears survive to older ages.
5. Regression and correlation. The idea of prediction is introduced through prediction of weight (very difficult to measure in a big bear) from chest girth (ok if the bear is anaesthetised). Outliers are introduced with possible reasons for the departure from the pattern.
6. Central limit theorem. The idea of random sampling is introduced by displaying samples of scaled bears and an xbar-bear. A histogram is evolved from the mean of each sample and finally a curve is fitted.
7. Estimating with confidence. Following on from 6 above random sampling shows where most xbars fall within a population distribution and how well sample means approximate the population mean. Decision making is required in another routine which introduces deliberately extreme sample means randomly. Estimating for unknown or hypothetical populations is accomplished by introducing a sample of ‘drop bears’.
8. Hypothesis tests are introduced and developed using random samples similarly to above with some deliberate biased samples introduced at random.